

Case study on IBM Employee attrition dataset using Pig

Data description: The dataset ibm.csv of IBM consists the information of the employee's which has some of the fields like

Age	HourlyRate
Attrition	Education
Business Travel	Job Role
Department	OverTime
Gender	Monthly Income
Employee Number	
Income	
Years_in_Current_role	
Years_since_Last_promotion	

Problem statements:

- 1) Find out the employee number and department of those employees who does overtime?
- 2) Find out the list of employee whose income is more than the average income of all the employee's present in the same department?
- 3) Count the no of employees who can be awarded as best employee of the year (hint: the concerned person must have rating above 3 and he/she must be working from past 5 years in the organization)
- 4) Display the number of employees who are in the queue of promotion
- 5) Display the list of employees who are exempted from the HRA
- 6) List the number of attrit employees
- 7) List out the department's where the female employees are lesser in number